

Relationship between edible fraction and physical attributes of *elabatu* (*Solanum melongena* l. var. *insanum*) fruit

M.P. Charithangi^{1*} and D.A.N. Dharmasena²

¹ Post Graduate Institute of Agriculture, University of Peradeniya, Sri Lanka.

² Department of Agricultural Engineering, Faculty of Agriculture, University of Peradeniya, Sri Lanka.

Received on June 27, 2023; Accepted on July 22, 2023

Abstract

Solanum melongena var. *insanum* known as *Elabatu* is a popular vegetable and a functional food consumed in Sri Lanka and it has two main cultivars bearing small and large size fruits. The fruit comprises of both edible and inedible portions and the consumer pays for both at the time of purchase. Therefore, this study was conducted to fit multiple regression models for size attributes and edible portion (EP) of *Elabatu* fruits as it is important to breeders as well as consumers. Small fruits with green stripes (Cultivar 1) and large fruits with white/purple stripes (Cultivar 2) were used in this study. Mass, smallest diameter, largest diameter, largest to smallest diameter ratio (LS ratio), percentage of edible portion to total fruit mass (EP %) and the mass of edible portion were measured. A sample of 195 fruits was used from each cultivar and data were fitted to multiple regression models. The best fitting models selected based on R^2 and least AIC values for cultivar 1 and cultivar 2 were $Y = \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 - \epsilon$ and $Y = \beta_3 X_3 - \beta_5 X_5 + \epsilon$, respectively.

Citation:

Charithangi, M., and Dharmasena, D. (2023). Relationship between edible fraction and physical attributes of *elabatu* (*Solanum melongena* l. Var. *insanum*) fruit. *Journal of Dry Zone Agriculture*, 9(1), 66-80.

*Corresponding author

Postal address: Department of Export Agriculture, Faculty of Animal Science and Export Agriculture, UvaWellassa University of Sri Lanka, Passara Road, Badulla..

Email: charithangi30@gmail.com



This article is published under the terms of a Creative Commons license: Attribution 4.0 International (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>).

Further, results revealed that the EP displays significant positive correlations ($p < 0.0001$) with LD, Mass, % of EP, and SD in both cultivars. In contrast to that, LS Ratio shows a significant negative correlation in Cultivar 1 ($r = -0.207$, $p = 0.0037$) and a significant positive correlation in Cultivar 2 ($r = 0.232$, $p = 0.0009$). Notably, Cultivar 2 shows stronger overall correlation coefficients.

Keywords: *Elabatu*; Mass of edible portion; Physical properties; Regression model fitting

Introduction

Solanum genus belongs to the family Solanaceae and it is the most economical and largest genus in this family and it comprises of 1 400 species worldwide (Chapman, 2019). *Solanum melongena* L. or cultivated brinjal / eggplant is economically important in many parts of the Asia (Oladosu *et al.*, 2021). China is in the forefront as the dominant producer, accounting for 62% of the total production in 2017 (Nadeeshani *et al.*, 2021). *Solanum melongena* var. *insanum* or *Elabatu* is considered as the primary gene pool of this *S. melongena* (Bohs, 2005) and the var. *insanum* is also called as wild brinjal.

Elabatu is grown in many parts of Sri Lanka which is widely called as “*Thalanabatu*” and used for indigenous medicine. Although roots are heavily used for medicinal purposes, the fruits are also used and they have significant analgesic, antipyretic, anti-inflammatory and Central Nervous System (CNS) depressant properties. CNS depressants are used to treat insomnia, anxiety and panic attacks (Ranil *et al.*, 2017). *Elabatu* is a highly cross-pollinated small shrub growing up to 1 m in height and its prickles are short, straight, and very sharp. Leaf grows up to 20 cm in length and 4 - 12 cm in width. (Aubriot and Knapp, 2022; Subasinghe *et al.*, 2003).

Fruits are mostly oval shaped with green, white or purple colour peel with stripes and turns into yellow when ripe (Knapp *et al.*, 2013). Discoid seeds in white flesh are seen in the middle of the fruit hence it is not possible to eat the entire fruit in traditionally grown cultivars. Although a modern hybrid is available with a completely edible fruit, majority of consumers prefer wild types due to their medicinal value. Middle white fleshy part should be removed before preparing a curry. However, consumer pays for the edible as well as non-edible portions at purchase. A relationship between physical dimensions and

edible portion of the fruit is important for breeders and also economically beneficial to the consumer.

Therefore, the objective of this study was to develop a mathematical relationship between the physical properties of the fruit and its edible portion for the two traditionally grown *Elabatu* cultivars in Sri Lanka. This information will facilitate the identification of superior fruit cultivar with a substantial edible mass while providing economic benefits to the consumer and valuable insights to the breeder.

Materials and methods

Sample collection

Two major *Elabatu* cultivars grown in Sri Lanka namely, Wild type/ Small fruit with green stripes (Cultivar 1) and White colour large fruit with green stripes (Cultivar 2) were used for this study (Fig. 1). Mature *Elabatu* fruits were purchased from the *Kuliyapitiya* weekly fair and safely transported to the laboratory. Sample size was selected as 200 fruits from each variety. However, it was reduced to 195 at the data collection stage due to some rejects.

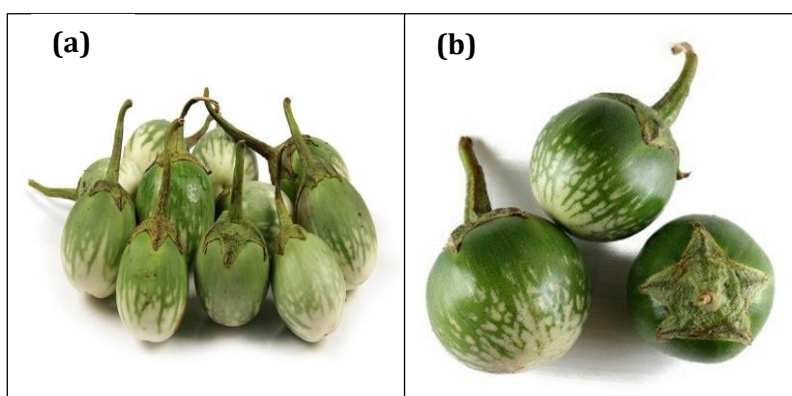


Fig. 1: Two major *Elabatu* cultivars grown in Sri Lanka: a - Wild type/ Small fruit with green stripes (Cultivar 1; b - White colour large fruit with green stripes (Cultivar 2)

Determination of linear measurements

The fruit size in terms of two linear measurements namely, Largest Diameter (LD) and Smallest Diameter (SD) (Fig. 2) were measured using a 0 – 300 mm (0

– 0.3 m) digital vernier caliper, with reading accuracy of 0.03 mm for 0 – 200 mm measuring range.

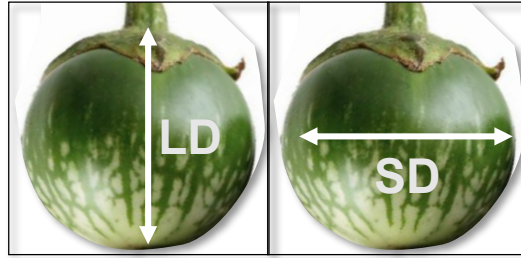


Fig. 2: Largest Diameter (LD) and Smallest Diameter (SD) of the fruit

Measurement of edible portion (EP) and percentage of edible portion (EP%)

Firstly, the mass of individual fruit was determined using a VIBRA LN CE series electronic balance (Maximum Capacity 4200 g (4.2 kg), Error 0.1 g (0.1×10^{-3} kg)). Then, fruits were cut into two portions along the center line and inedible parts in the middle portion were removed by using a spoon. Then, mass of the two portions were measured again by using the same balance. That was recorded as the mass of the edible portion of each *Elabatu* fruit. Then, percentage of edible portion was calculated by using following equation (Equation 1). In addition, individual fruit volumes were also recorded but it was dropped off in the analysis as it is dependent on the linear dimensions for this fruit shape.

$$\% \text{ of the edible fraction} = \frac{\text{mass of the edible fraction}}{\text{mass of the fruit}} \times 100 \quad (1)$$

Data analysis

The statistical analysis was performed using the SPSS statistics standard version 23. As the first step, descriptive statistics analysis was performed to describe the basic features of the data set. The mean and standard deviation statistical measures were used to compute further statistical testing for the present study.

As the next step; correlation coefficient analysis was done to measure the relationship strength of selected variables (Anderson, 1974). Moreover, a

multiple regression analysis was carried out to evaluate the effect of all the selected variables (LD, SD, LS Ratio, Mass and percentage of edible portion) to the Edible Portion. Further, gathered data from 195 fruits from each cultivar were fitted into multiple regression models and the best fit models were selected according to the R^2 value and least AIC (Akaike Information Criterion) values. Least AIC value was given for the best model (Sakamoto *et al.*, 1986).

In further analysis of this study, model checking techniques were applied after formulating the model. The raw materials of the model checking are the residuals defined as the differences between observed and fitted values (Anderson, 1974). Then Q- Q plots were graphed. Q- Q plot is one of the probability plots and diagnostic aid. Quantiles of the data versus quantiles of a distribution was shown in this Q- Q plot. In quantile-quantile (Q-Q) plots, two sets of points were compared: the observed data and the anticipated values derived from either a fitted model or a theoretical distribution.

By plotting the quantiles of one set against the quantiles of the other, they offer a visual means to evaluate the degree of alignment between the data and the expected distribution or model (Mackay and Jonathan, 2021). Then, multi collinearity detection was carried out to determine the impact of collinearity among the variables in regression models (Shrestha, 2020). Variance Inflation Factor (VIF) values were also obtained for present study. If $VIF(\widehat{\beta}_i) > 1$, multi collinearity is high and also 10 has been proposed as a cut off value (Alin, 2010).

Finally, Durbin- Watson test was used to test the auto correlation in the residuals of the regression analysis (Durbin and Watson, 1971). The Durbin-Watson statistic should always have a value between 0 and 4. Then, Breusch-Pagan (BP) test was carried out to determine whether there is a constant variance in the residuals (Breusch and Pagan, 1980). As the final step, predicted values were plotted with measured data using the model as validation graphs.

Results and discussion

Descriptive statistics analysis

Edible Portion (EP), Largest Diameter (LD), Smallest Diameter (SD), Largest Diameter / Smallest Diameter Ratio (LS Ratio), Fruit Mass and Mass of Edible

Portion to Total Fruit Percentage (% of EP) were the variables used in this analysis (Table 1).

Table 1: Descriptive statistics for cultivar 1 and 2

Variables	Mean		Standard Deviation		Sample Size (n)	
	Cultivar 1	Cultivar 2	Cultivar 1	Cultivar 2	Cultivar 1	Cultivar 2
EP (g)	3.123	15.370	0.889	6.863	195	195
LD (mm)	20.370	35.328	2.098	6.435	195	195
SD (mm)	20.288	0.978	2.179	0.110	195	195
LS Ratio	1.008	25.061	0.080	10.270	195	195
Mass (g)	5.272	61.158	1.353	8.220	195	195
% of EP	59.580	36.177	10.718	5.492	195	195

Correlation matrix for varieties

Hypothesis for matrix (Table 2) was $-1 < r > 1$ with 95 % confidence level ($P = 0.05$). H_0 was defined as there is no relationship between the variables and H_A was considered as there is a relationship between the variables.

Table 2: Correlation matrix for Cultivar 1

Variables	EP	LD	LS Ratio	Mass	% of EP	SD
EP (g)	1.000	0.613 (****)	-0.207 (**)	0.746 (****)	0.524 (****)	0.748 (****)
LD (mm)	0.613 (****)	1.000	0.348 (****)	0.729 (****)	-0.060 (ns)	0.686 (****)
LS Ratio	-0.207 (**)	0.348 (****)	1.000	-0.207 (**)	-0.112 (ns)	-0.440 (****)
Mass (g)	0.746 (****)	0.729 (****)	-0.207 (**)	1.000	-0.121 (ns)	0.873 (****)
% of EP	0.524 (****)	-0.060 (ns)	-0.112 (ns)	-0.121 (ns)	1.000	0.013 (ns)
SD (mm)	0.748 (****)	0.686 (****)	-0.440 (****)	0.873 (****)	0.013 (ns)	1.000

($P = 0.05$, ns $P > 0.05$, ** $P \leq 0.01$, **** $P \leq 0.0001$)

According to the results of this study; null hypothesis was rejected, and these results revealed that the value of r in Smallest Diameter was fairly high (0.87) with Mass of the fruit (Table 2). It was found that, relationship between Mass with the Smallest Diameter was strong and positively correlated with each other ($P < 0.0001$). Largest Diameter and Percentage of Edible Portion to total fruit (% EP) showed a significant moderate relationship with the mass of Edible Portion (EP) but those were positively correlated. LS ratio showed a weak relationship with edible portion and it was negatively correlated.

Table 3: Correlation matrix for Cultivar 2

Variables	EP	LD	LS Ratio	Mass	% of EP	SD
EP (g)	1.000	0.847 (****)	0.232 (***)	0.937 (****)	0.377 (****)	0.845 (****)
LD (mm)	0.847 (****)	1.000	0.555 (****)	0.924 (****)	-0.011 (ns)	0.777 (****)
LS Ratio	0.232 (***)	0.555 (****)	1.000	0.280 (****)	-0.089 (ns)	-0.086 (ns)
Mass (g)	0.937 (****)	0.924 (****)	0.281 (****)	1.000	0.051 (ns)	-0.899 (****)
% of EP	0.377 (****)	-0.011 (ns)	-0.089 (ns)	0.051 (ns)	1.000	0.051 (ns)
SD (mm)	0.845 (****)	0.777 (****)	-0.086 (ns)	0.899 (****)	0.051 (ns)	1.000

($P = 0.05$, ns $P > 0.05$, ** $P \leq 0.01$, **** $P \leq 0.0001$)

Similar to the cultivar 1, null hypothesis was rejected and these results revealed that, Mass, Largest Diameter and Smallest Diameter of the fruit showed a strong relationship (0.85-0.9) with mass of Edible Portion of the fruit (EP) and positively correlated with each other (Table 3). LS ratio and Percentage of edible portion (% EP) showed a weak relationship.

Regression analysis

The regression model used for this analysis is given as follows (Equation 2).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (2)$$

Where: $Y = EP$; $X_1 = \text{Smallest Diameter}$; $X_2 = \text{Largest Diameter}$; $X_3 = \text{Mass}$; $X_4 = \text{Percentage of Edible Portion}$; $X_5 = \text{LS Ratio}$; $\varepsilon = \text{Residual Error}$ and $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ and β_5 are regression coefficients. It was found that the linear relationship was stronger than other models tried.

Model selection

A summary of AIC values for each cultivar have been represented in Table 4.

Table 4: AIC values of cultivar 1 and cultivar 2

Cultivar 1		
Steps	Model	AIC Value
01	EP ~ LD + SD + Mass + Percentage of Edible Portion + LS Ratio	-604.64
02	EP ~ SD + Mass + Percentage of Edible Portion + LS Ratio	-605.94
Cultivar 2		
Steps	Model	AIC Value
01	EP ~ LD + SD + Mass + Percentage of Edible Portion + LS Ratio	-918.79
02	EP ~ SD + Mass + Percentage of Edible Portion + LS Ratio	-919.50
03	EP ~ Mass + Percentage of Edible Portion + LS Ratio	-920.92
04	EP ~ Mass + LS Ratio	-921.41

From among the models tested, the EP as a function of SD + Mass + Percentage of Edible Portion + LS Ratio model was considered as the best fitting model for cultivar 1. Conversely, the EP as a function of Mass + LS Ratio model was the best fit model for cultivar 2, as those two models had the minimum AIC value compared to the other models.

Analysis of variance (ANOVA) tables

Goodness of fit and parameter estimation of the best fitting model for each cultivar was done in this step (Table 5 and Table 6).

Table 5: Goodness of fit and parameter estimates of the best fitting model for Cultivar 1

Parameter	Estimate	Standard Error	Pr(> t)
(Intercept)	-4.804	0.412	< 0.0001
SD	0.081	0.017	< 0.0001
Mass	0.436	0.026	< 0.0001
LS Ratio	0.953	0.228	< 0.0001
% of EP	0.051	0.001	< 0.0001

(Residual standard error: 0.2088, multiple r- squared: 0.946, adjusted r- squared: 0.9449)

Model obtained with four variables for predicting the Edible Portion (EP) of the Cultivar 1 is as follows:

$$\text{EP} = 0.08 \text{ SD} + 0.43 \text{ Mass} + 0.95 \text{ LS Ratio} + 0.05 (\% \text{ of Edible Portion}) - 4.80$$

Table 6: Goodness of fit and parameter estimates of the best fitting model for Cultivar 2

Parameter	Estimate	Standard Error	Pr(> t)
(Intercept)	2.491	1.503	0.099
LS Ratio	-0.135	0.069	0.051
Mass	0.704	0.043	< 0.0001

(Residual standard error: 2.392, multiple r-squared: 0.8798, adjusted r-squared: 0.8785)

Model obtained with two variables for predicting the mass of Edible Portion (EP) of Cultivar 2 is given below.

$$\text{EP} = 0.70 \text{ Mass} - 0.13 \text{ LS Ratio} + 2.49$$

Further analysis

Fig. 3 shows the data distribution pattern of the residuals and fitted values of the Cultivar 1 and 2 respectively.

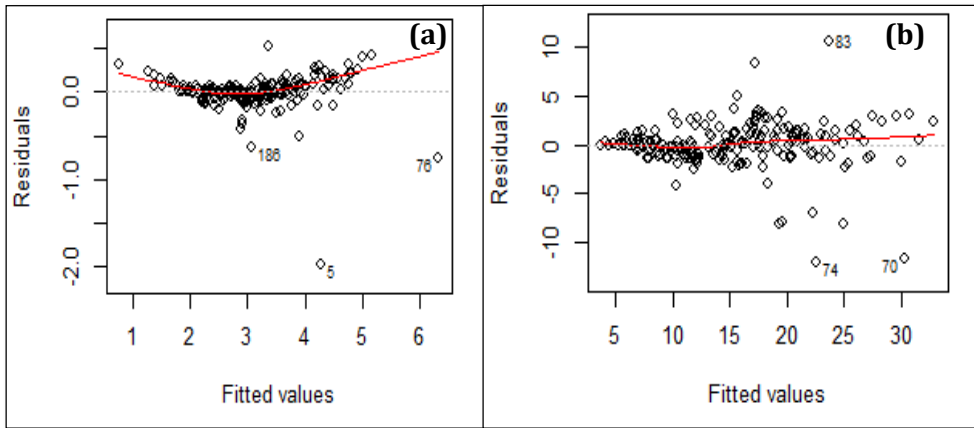


Fig. 3: Fitted values Vs Residuals for Cultivar 1 (a) and Cultivar 2 (b)

Fig. 4 shows the data distribution pattern of the standard residuals with theoretical quantiles of the Cultivar 1 and 2 respectively.

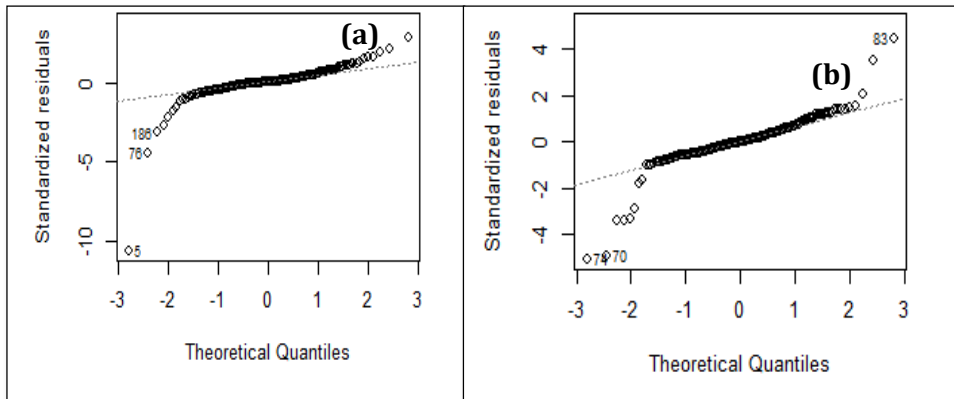


Fig. 4: Theoretical quantiles Vs Standard residuals for Cultivar 1 (a) and Cultivar (b)

Multicollinearity detection for the multiple regressions

Normally, values of VIF that exceed 10 are often regarded as indicating multicollinearity. Moderate VIF values obtained for these regression models (6.198 for one cultivar and 6.848 for the other) can be attributed to the notable high correlation between Mass and both LD and SD. Such relatively high correlation among independent variables can lead to collinearity issues, potentially inflating the VIF values. Although the VIF values do not surpass the commonly

accepted threshold of 10, it is important to be cautious in interpreting the results and discerning the individual impacts of Mass, LD, and SD on the prediction of Edible Portion (EP) as a mass. The presence of interrelatedness among these variables suggests the need for thoughtful consideration when interpreting the coefficients and their respective contributions to the model's predictions.

Durbin- Watson test

Following test results were obtained for the two cultivars. The DW of 2.0639 and a P value of 0.6774 for the cultivar 1 and the DW of 1.3139 and p-value ≤ 0.0001 for the cultivar 2. These results revealed that true autocorrelation is not zero for the present study (H_0 : Residuals are uncorrelated (Autocorrelation is zero); H_1 : Residuals are correlated (Autocorrelation is not zero)).

Breusch-Pagan test

The following hypotheses were used for the test.

H_0 : the variance is constant in the data set

H_A : the variance is not constant in the data set

BP value obtained for the variety 1 is 117.39 with a p-value < 0.001 and the BP value for the variety 2 is 79.118 with a p-value < 0.001 respectively. Therefore, there is no evidence that the errors have non-constant variance in this data set.

Based on all the above analysis tools, the best regression model to calculate the edible portion of *Elabatu* for cultivar 1 is given below (Equation 3).

$$Y = \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 - \varepsilon \quad (3)$$

(R^2 0.9448)

Where: X_1 = smallest diameter of the fruit; X_3 = mass of the fruit; X_4 = percentage of the edible portion; X_5 = largest diameter to smallest diameter ratio; Y = mass of edible portion; ε = error and regression coefficients and the error are: $\beta_1 = 0.08$; $\beta_3 = 0.43$; $\beta_4 = 0.05$; $\beta_5 = 0.95$; $\varepsilon = 4.80$

The best regression model to calculate the edible portion of Cultivar 2 is given below (Equation 4).

$$Y = \beta_3 X_3 - \beta_5 X_5 + \varepsilon \quad (4)$$

(R^2 0.8785)

Where: X_3 = mass of the fruit; X_5 = largest diameter to smallest diameter ratio; Y = edible portion; ϵ = error and the regression coefficients; $\beta_3 = 0.70$ and $\beta_5 = -0.13$ with and error (ϵ) of 2.49.

Finally, actual values were fitted to the predicted values using model validation graphs. The predicted and observed values for cultivar 1 and 2 were plotted and shown in Figs. 5 and 6 respectively. It was found that the patterns of predicted and observed values are more or less the same and most points are overlapped. This indicates that the model has a high accuracy for describing the physical parameters of the fruit.

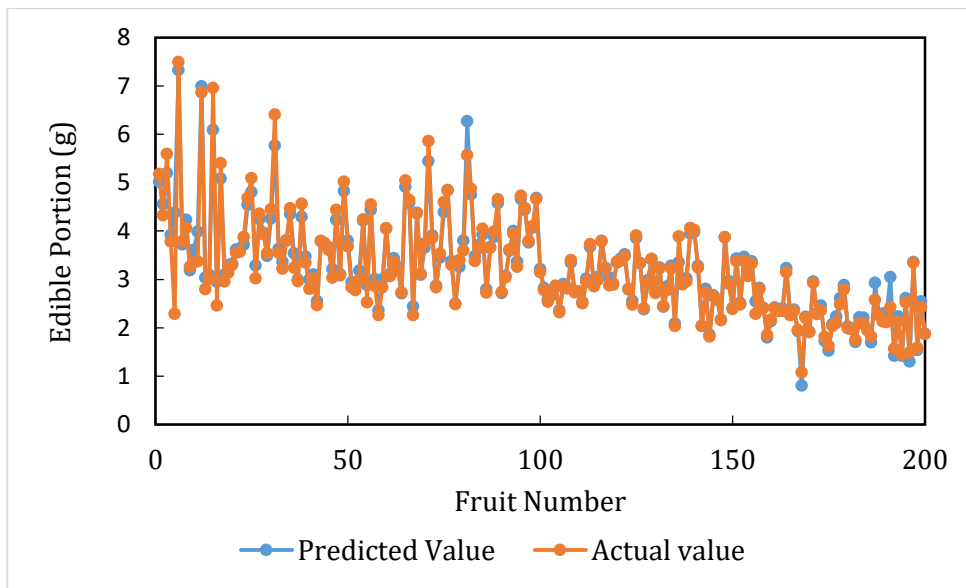


Fig. 5: model validation graph for Cultivar 1

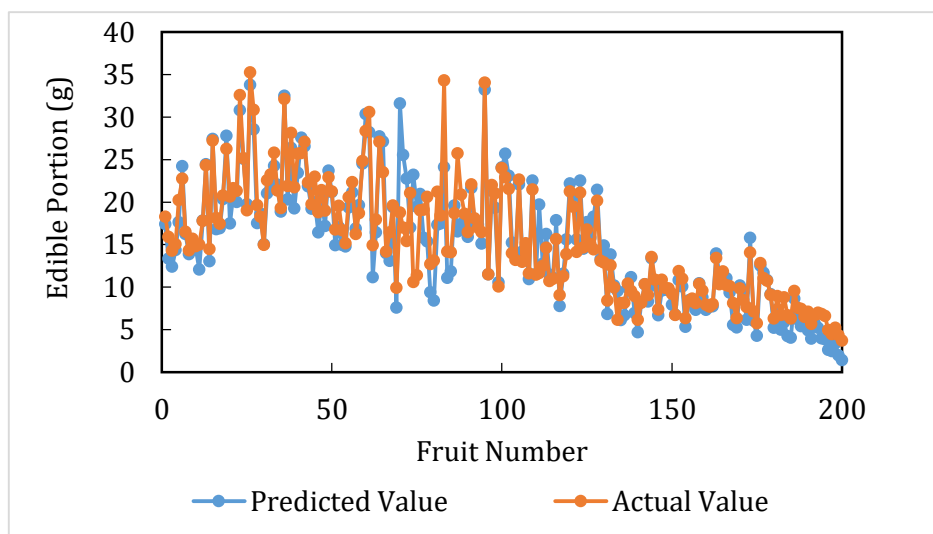


Fig. 6: Model validation graph for Cultivar 2

Conclusion

The study reveals that the best model for predicting the edible portion of Cultivar 1 is $Y = \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 - \epsilon$ and can be estimated based on the physical properties; Smallest Diameter (SD), Largest Diameter to Smallest Diameter Ratio (LS Ratio), Mass of the fruit and percentage of edible portion (% EP). Based on the results, the mass of edible portion is directly proportional to the fruit mass, smallest diameter, largest diameter to smallest diameter ratio. Therefore, largest, oval shape fruits are better than small, oblong fruits.

The best fitted model for Cultivar 2 can be estimated using the two physical properties; mass of the fruit and the largest diameter to smallest diameter ratio using the $Y = \beta_3 X_3 - \beta_5 X_5 + \epsilon$ model. For cultivar 2, edible portion is directly proportional to fruit mass and the mass of edible portion declines with the increase of the largest diameter to smallest diameter ratio. When buying the green colour, white striped large fruit (cultivar 2), the most economical way is to select the large fruit.

Acknowledgement

Authors would like to acknowledge the University College of Kuliapitiya (UCK) for providing laboratory facilities to conduct this research conveniently.

Declaration of conflict of interest

Authors have no conflict of interest to declare.

References

- Alin, A. (2010). Multicollinearity. *Wiley interdisciplinary reviews: computational statistics*, 2(3), 370-374.
- Anderson, T.W. 1974, *Introductory statistical analysis*. Houghton Mifflin Company, Massachusetts. pp
- Aubriot, X., and Knapp, S. (2022). A revision of the “*Spiny solanums*” of Tropical Asia (*Solanum*, the *Leptostemonum* Clade, *Solanaceae*). *PhytoKeys*, 198, 1.
- Bohs, L. (2005). Major clades in *Solanum* based on *ndhF* sequence data. *Monographs in Systematic Botany*, 104, 27.
- Breusch, T. S., and Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The review of economic studies*, 47(1), 239-253.
- Chapman, M.A., 2019. *Introduction: The importance of eggplant*. In *The Eggplant Genome*; Springer: Cham, Switzerland, pp. 1–10.
- Durbin, J., and Watson, G. S. (1971). Testing for serial correlation in least squares regression. III. *Biometrika*, 58(1), 1-19.
- Knapp, S., Vorontsova, M. S., and Prohens, J. (2013). Wild relatives of the eggplant (*Solanum melongena* L.: *Solanaceae*): new understanding of species names in a complex group. *PloS one*, 8(2), e57039.
- Mackay, E., and Jonathan, P. (2021). Sampling properties and empirical estimates of extreme events. *Ocean Engineering*, 239, 109791.
- Nadeeshani, H., Samarasinghe, G., Wimalasiri, S., Silva, R., Hunter, D., and Madhujith, T. (2021). Comparative analysis of the nutritional profiles of

selected *Solanum* species grown in Sri Lanka. *Journal of Food Composition and Analysis*, 99, 103847.

Oladosu, Y., Rafii, M. Y., Arolu, F., Chukwu, S. C., Salisu, M. A., Olaniyan, B. A., ... and Muftaudeen, T. K. (2021). Genetic diversity and utilization of cultivated eggplant germplasm in varietal improvement. *Plants*, 10(8), 1714.

Ranil, R. H., Prohens, J., Aubriot, X., Niran, H. M. L., Plazas, M., Fonseka, R. M., ... and Knapp, S. (2017). *Solanum insanum* L.(subgenus *Leptostemonum* Bitter, Solanaceae), the neglected wild progenitor of eggplant (*S. melongena* L.): a review of taxonomy, characteristics and uses aimed at its enhancement for improved eggplant breeding. *Genetic resources and crop evolution*, 64, 1707-1722.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, 81(10.5555), 26853.

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39-42.

Subasinghe, S., Senaratne, R., and Weerasinghe, W. I. L., 2003. Development of effective propagation techniques for Elabatu. <http://ir.lib.ruh.ac.lk/xmlui/handle/iruor/256>, Accessed June 15 2023.